

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	0	classify\$4 same automatic\$5 same word\$4 same annotation\$1 same concordance\$1	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2005/02/16 11:18
L2	3	assign\$4 same normaliz\$5 same weight\$1 same automatic\$5 same annotation\$1	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2005/02/16 11:19
L3	27	generat\$4 same automatic\$5 same (database\$1 or db\$1 or (data adj base\$1)) same design\$4 same schema\$1	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2005/02/16 11:21
L4	0	L3 and L2	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2005/02/16 11:21

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	15	("5884302", "6108666", "6233571", "6272242", "649821", "6694313"). pn.	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2005/02/09 17:33
L2	0	L1 and (annotat\$5 same schema\$1)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2005/02/09 17:34
L3	0	L1 and (annotat\$5 and schema\$1)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2005/02/09 18:08
L4	1	L1 and annotat\$5	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2005/02/09 17:34
L5	0	L1 and (schema\$1)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2005/02/09 18:09
L6	619	natural\$4 same langu\$7 same quer\$3	USPAT	OR	ON	2005/02/09 18:09
L7	1444	natural\$4 same langu\$7 same quer\$3	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2005/02/09 18:11
L8	402	L7 and (quer\$3 same translat\$4)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2005/02/09 18:11
L9	259	L8 and (generat\$4 same (database\$1 db\$1 or (data adj base\$1)))	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2005/02/09 18:12

L10	35	L9 and (design\$4 same schema\$1)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2005/02/09 18:13
-----	----	-----------------------------------	---	----	----	------------------



IEEE Xplore®

RELEASE 1.8

Welcome
United States Patent and Trademark Office


[Help](#) | [FAQ](#) | [Terms](#) | [IEEE Peer Review](#)
[Quick Links](#)

Welcome to IEEE Xplore®

- ☐ Home
- ☐ What Can I Access?
- ☐ Log-out

Tables of Contents

- ☐ Journals & Magazines
- ☐ Conference Proceedings
- ☐ Standards

Search

- ☐ By Author
- ☐ Basic
- ☐ Advanced
- ☐ CrossRef

Member Services

- ☐ Join IEEE
- ☐ Establish IEEE Web Account
- ☐ Access the IEEE Member Digital Library

IEEE Enterprise

- ☐ Access the IEEE Enterprise File Cabinet

Print Format

Your search matched **0** of **1128145** documents.

A maximum of **500** results are displayed, **15** to a page, sorted by **Relevance** in **Descending** order.

Refine This Search:

You may refine your search by editing the current search expression or entering a new one in the text box.

☐ Check to search within this result set

Results Key:

JNL = Journal or Magazine **CNF** = Conference **STD** = Standard

Results:

No documents matched your query.

[\[Abstract\]](#) [\[PDF Full-Text \(816 KB\)\]](#) IEEE JNL

6 Automatic Linguistic Indexing of Pictures by a statistical modeling approach

Jia Li; Wang, J.Z.;

Pattern Analysis and Machine Intelligence, IEEE Transactions on , Volume: 25 , Issue: 9 , Sept. 2003

Pages:1075 - 1088

[\[Abstract\]](#) [\[PDF Full-Text \(3240 KB\)\]](#) IEEE JNL

7 Automatic ontology-based knowledge extraction from Web documents

Alani, H.; Sanghee Kim; Millard, D.E.; Weal, M.J.; Hall, W.; Lewis, P.H.; Shadbolt, N.R.;

Intelligent Systems, IEEE [see also IEEE Expert] , Volume: 18 , Issue: 1 , Jan.-Feb. 2003

Pages:14 - 21

[\[Abstract\]](#) [\[PDF Full-Text \(2527 KB\)\]](#) IEEE JNL

8 Musical genre classification of audio signals

Tzanetakis, G.; Cook, P.;

Speech and Audio Processing, IEEE Transactions on , Volume: 10 , Issue: 5 , July 2002

Pages:293 - 302

[\[Abstract\]](#) [\[PDF Full-Text \(279 KB\)\]](#) IEEE JNL

9 An experimental study on content-based image classification for satellite image databases

Holowczak, R.D.; Artigas, F.J.; Soon Ae Chun; June-Suh Cho; Stone, H.S.;

Geoscience and Remote Sensing, IEEE Transactions on , Volume: 40 , Issue: 6 , June 2002

Pages:1338 - 1347

[\[Abstract\]](#) [\[PDF Full-Text \(628 KB\)\]](#) IEEE JNL

10 A practical approach to programming with assertions

Rosenblum, D.S.;

Software Engineering, IEEE Transactions on , Volume: 21 , Issue: 1 , Jan. 1995

Pages:19 - 31

[\[Abstract\]](#) [\[PDF Full-Text \(1200 KB\)\]](#) IEEE JNL

11 Detecting Tactics Patterns for Archiving Tennis Video Clips

Jenny R. Wang, Nandan Parameswaran;

Multimedia Software Engineering, 2004. Proceedings. IEEE Sixth International Symposium on , 13-15 Dec. 2004

Pages:186 - 192

[\[Abstract\]](#) [\[PDF Full-Text \(208 KB\)\]](#) IEEE CNF

12 eMeeting: a multimedia application for interactive meeting and seminar

Wing Ho Leung; Tsuhan Chen; Hendriks, F.; Xiping Wang; Zon-Yin Shae;

Global Telecommunications Conference, 2002. GLOBECOM '02. IEEE , Volume: 3 , 17-21 Nov. 2002

Pages:2994 - 2998 vol.3

[\[Abstract\]](#) [\[PDF Full-Text \(432 KB\)\]](#) [IEEE CNF](#)

13 Generating fuzzy semantic metadata describing spatial relations from images using the R-histogram

Yuhang Wang; Makedon, F.; Ford, J.; Shen, L.; Goldin, D.;

Digital Libraries, 2004. Proceedings of the 2004 Joint ACM/IEEE Conference on , 7-11 June 2004

Pages:202 - 211

[\[Abstract\]](#) [\[PDF Full-Text \(364 KB\)\]](#) [IEEE CNF](#)

14 Semantic video classification and feature subset selection under context and concept uncertainty

Fan, J.; Luo, H.; Xiao, J.; Wu, L.;

Digital Libraries, 2004. Proceedings of the 2004 Joint ACM/IEEE Conference on , 7-11 June 2004

Pages:192 - 201

[\[Abstract\]](#) [\[PDF Full-Text \(430 KB\)\]](#) [IEEE CNF](#)

15 SVM-based audio classification for instructional video analysis

Ying Li; Chitra Dorai;

Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on , Volume: 5 , 17-21 May 2004

Pages:V - 897-900 vol.5

[\[Abstract\]](#) [\[PDF Full-Text \(226 KB\)\]](#) [IEEE CNF](#)

[1](#) [2](#) [3](#) [Next](#)



IEEE Xplore®

RELEASE 1.8

Welcome
United States Patent and Trademark Office


[Help](#) [FAQ](#) [Terms](#) [IEEE Peer Review](#)
[Quick Links](#)

Welcome to IEEE Xplore®

- ☐ Home
- ☐ What Can I Access?
- ☐ Log-out

Tables of Contents

- ☐ Journals & Magazines
- ☐ Conference Proceedings
- ☐ Standards

Search

- ☐ By Author
- ☐ Basic
- ☐ Advanced
- ☐ CrossRef

Member Services

- ☐ Join IEEE
- ☐ Establish IEEE Web Account
- ☐ Access the IEEE Member Digital Library

IEEE Enterprise

- ☐ Access the IEEE Enterprise File Cabinet

Print Format

Your search matched **36** of **1128145** documents.

A maximum of **500** results are displayed, **15** to a page, sorted by **Relevance** in **Descending** order.

Refine This Search:

You may refine your search by editing the current search expression or entering a new one in the text box.

☐ Check to search within this result set

Results Key:

JNL = Journal or Magazine **CNF** = Conference **STD** = Standard

16 Semantics-oriented language understanding with automatic adaptability

Feng, D.; Hovy, E.;

Natural Language Processing and Knowledge Engineering, 2003. Proceedings. 2003 International Conference on , 26-29 Oct. 2003

Pages:33 - 39

[\[Abstract\]](#) [\[PDF Full-Text \(431 KB\)\]](#) IEEE CNF

17 Automatic snoring signal analysis in sleep studies

Jane, R.; Fiza, J.A.; Sola-Soler, J.; Blanch, S.; Artis, P.; Morera, J.;

Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE , Volume: 1 , 17-21 Sept. 2003

Pages:366 - 369 Vol.1

[\[Abstract\]](#) [\[PDF Full-Text \(452 KB\)\]](#) IEEE CNF

18 Image classification using multimedia knowledge networks

Benitez, A.B.; Shih-Fu Chang;

Image Processing, 2003. Proceedings. 2003 International Conference on , Volume: 3 , 14-17 Sept. 2003

Pages:III - 613-16 vol.2

[\[Abstract\]](#) [\[PDF Full-Text \(395 KB\)\]](#) IEEE CNF

19 VideoAL: a novel end-to-end MPEG-7 video automatic labeling system

Ching-Yung Lin; Tseng, B.L.; Naphade, M.; Natsev, A.; Smith, J.R.;

Image Processing, 2003. Proceedings. 2003 International Conference on , Volume: 3 , 14-17 Sept. 2003

Pages:III - 53-6 vol.2

[\[Abstract\]](#) [\[PDF Full-Text \(516 KB\)\]](#) IEEE CNF

20 Evaluation strategies for automatic linguistic indexing of pictures

Wang, J.Z.; Jia Li; Sui Ching Lin;

Image Processing, 2003. Proceedings. 2003 International Conference on , Volume: 3 , 14-17 Sept. 2003
Pages:III - 617-20 vol.2

[\[Abstract\]](#) [\[PDF Full-Text \(409 KB\)\]](#) [IEEE CNF](#)

21 Expressive semantics for automatic annotation and retrieval of video streams

Del Bimbo, A.;

Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on , Volume: 2 , 30 July-2 Aug. 2000
Pages:671 - 674 vol.2

[\[Abstract\]](#) [\[PDF Full-Text \(552 KB\)\]](#) [IEEE CNF](#)

22 Integration of color, shape, and texture for image annotation and retrieval

Saber, E.; Tekalp, A.M.;

Image Processing, 1996. Proceedings., International Conference on , Volume: 3 , 16-19 Sept. 1996
Pages:851 - 854 vol.3

[\[Abstract\]](#) [\[PDF Full-Text \(544 KB\)\]](#) [IEEE CNF](#)

23 Highlight extraction in soccer videos

Assfalg, J.; Bertini, M.; Colombo, C.; Del Bimbo, A.; Nunziati, W.;

Image Analysis and Processing, 2003.Proceedings. 12th International Conference on , 17-19 Sept. 2003
Pages:498 - 503

[\[Abstract\]](#) [\[PDF Full-Text \(422 KB\)\]](#) [IEEE CNF](#)

24 Towards automated derivation of biological pathways using high-throughput biological data

Yu Chen; Trupti Joshi; Ying Xu; Dong Xu;

Bioinformatics and Bioengineering, 2003. Proceedings. Third IEEE Symposium on , 10-12 March 2003
Pages:18 - 25

[\[Abstract\]](#) [\[PDF Full-Text \(279 KB\)\]](#) [IEEE CNF](#)

25 Independent component analysis for understanding multimedia content

Kolenda, T.; Hansen, L.K.; Larsen, J.; Winther, O.;

Neural Networks for Signal Processing, 2002. Proceedings of the 2002 12th IEEE Workshop on , 4-6 Sept. 2002
Pages:757 - 766

[\[Abstract\]](#) [\[PDF Full-Text \(427 KB\)\]](#) [IEEE CNF](#)

26 Proceedings of 2002 International Conference on Machine Learning and Cybernetics (Cat.No.02EX583)

Machine Learning and Cybernetics, 2002. Proceedings. 2002 International Conference on , Volume: 2 , 4-5 Nov. 2002

[\[Abstract\]](#) [\[PDF Full-Text \(1409 KB\)\]](#) [IEEE CNF](#)

27 Colour-based object recognition for video annotation*Koubaroulis, D.; Matas, J.; Kittler, J.;*

Pattern Recognition, 2002. Proceedings. 16th International Conference on , Volume: 2 , 11-15 Aug. 2002

Pages:1069 - 1072 vol.2

[\[Abstract\]](#) [\[PDF Full-Text \(446 KB\)\]](#) [IEEE CNF](#)**28 Automatic sports classification***Messer, K.; Christmas, W.; Kittler, J.;*

Pattern Recognition, 2002. Proceedings. 16th International Conference on , Volume: 2 , 11-15 Aug. 2002

Pages:1005 - 1008 vol.2

[\[Abstract\]](#) [\[PDF Full-Text \(366 KB\)\]](#) [IEEE CNF](#)**29 Extraction of TV highlights using multimedia features***Dagtas, S.; Abdel-Mottaleb, M.;*

Multimedia Signal Processing, 2001 IEEE Fourth Workshop on , 3-5 Oct. 2001

Pages:91 - 96

[\[Abstract\]](#) [\[PDF Full-Text \(412 KB\)\]](#) [IEEE CNF](#)**30 Augmented edit distance based temporal contiguity analysis for improved videotext recognition***Aradhye, H.; Dorai, C.;*

Multimedia Signal Processing, 2001 IEEE Fourth Workshop on , 3-5 Oct. 2001

Pages:275 - 280

[\[Abstract\]](#) [\[PDF Full-Text \(504 KB\)\]](#) [IEEE CNF](#)[Prev](#) [1](#) [2](#) [3](#) [Next](#)

segmentation*Tzanetakis, G.; Cook, F.;*

EUROMICRO Conference, 1999. Proceedings. 25th , Volume: 2 , 8-10 Sept. 1999

Pages:61 - 67 vol.2

[\[Abstract\]](#) · [\[PDF Full-Text \(116 KB\)\]](#) IEEE CNF

36 ClassX: a browsing tool for protein sequence megaclassifications*Harris, N.L.; States, D.J.; Hunter, L.;*

System Sciences, 1993, Proceeding of the Twenty-Sixth Hawaii International Conference on , Volume: i , 5-8 Jan 1993

Pages:554 - 563 vol.1

[\[Abstract\]](#) [\[PDF Full-Text \(916 KB\)\]](#) IEEE CNF

[Prev](#) [1](#) [2](#) [3](#)

[Home](#) | [Log-out](#) | [Journals](#) | [Conference Proceedings](#) | [Standards](#) | [Search by Author](#) | [Basic Search](#) | [Advanced Search](#) | [Join IEEE](#) | [Web Account](#) | [New this week](#) | [O](#)
[Linking Information](#) | [Your Feedback](#) | [Technical Support](#) | [Email Alerting](#) | [No Robots Please](#) | [Release Notes](#) | [IEEE Online Publications](#) | [Help](#) | [FAQ](#) | [Terms](#) | [Back to](#)

Copyright © 2004 IEEE — All rights reserved



US Patent & Trademark Office

[Subscribe \(Full Service\)](#) [Register \(Limited Service, Free\)](#) [Login](#)

 Search: ☒ The ACM Digital Library ☐ The Guide


[Feedback](#) [Report a problem](#) [Satisfaction survey](#)

Terms used

classif and automatic and annotation and concordanc and normaliz and weight and word

Found 20,783 of

150,138

 Sort results
by

 Display
results


[Save results to a Binder](#)

[Search Tips](#)
☐ Open results in a new window

[Try an Advanced Search](#)
[Try this search in The ACM Guide](#)

Results 1 - 20 of 200

 Result page: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [next](#)

Best 200 shown

 Relevance scale ☐ ☐ ☐ ☐ ☐

1 [Models of translational equivalence among words](#)

I. Dan Melamed

June 2000 **Computational Linguistics**, Volume 26 Issue 2

Full text available:


 Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#)
[Publisher Site](#)

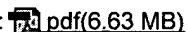
Parallel texts (bitexts) have properties that distinguish them from other kinds of parallel data. First, most words translate to only one other word. Second, bitext correspondence is typically only partial---many words in each text have no clear equivalent in the other text. This article presents methods for biasing statistical translation models to reflect these properties. Evaluation with respect to independent human judgments has confirmed that translation models biased in this fashion are si ...

2 [Curriculum 68: Recommendations for academic programs in computer science: a report of the ACM curriculum committee on computer science](#)

William F. Atchison, Samuel D. Conte, John W. Hamblen, Thomas E. Hull, Thomas A. Keenan, William B. Kehl, Edward J. McCluskey, Silvio O. Navarro, Werner C. Rheinboldt, Earl J. Schweppe, William Viavant, David M. Young

March 1968 **Communications of the ACM**, Volume 11 Issue 3

Full text available:


 Additional Information: [full citation](#), [references](#), [citations](#)

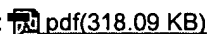
Keywords: computer science academic programs, computer science bibliographies, computer science courses, computer science curriculum, computer science education, computer science graduate programs, computer science undergraduate programs

3 [Multimedia data mining: Automatic image annotation and retrieval using subspace clustering algorithm](#)

Lei Wang, Li Liu, Latifur Khan

November 2004 **Proceedings of the 2nd ACM international workshop on Multimedia databases**

Full text available:


 Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

The development of technology generates huge amounts of non-textual information, such as images. An efficient image annotation and retrieval system is highly desired. Clustering algorithms make it possible to represent visual features of images with finite symbols. Based on this, many statistical models, which analyze correspondence between visual features and words and discover hidden semantics, have been published. These models improve the

annotation and retrieval of large image databases. ...

Keywords: EM, SVD, blob tokens, features, similarity, subspace clustering

4 Word sense disambiguation of adjectives using probabilistic networks

Gerald Chao, Michael G. Dyer

July 2000 **Proceedings of the 17th conference on Computational linguistics - Volume 1**

Full text available:  [pdf\(715.26 KB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#)

In this paper, word sense disambiguation (WSD) accuracy achievable by a probabilistic classifier, using very minimal training sets, is investigated. We made the assumption that there are *no* tagged corpora available and identified what information, needed by an accurate WSD system, can and cannot be automatically obtained. The lesson learned can then be used to focus on what knowledge needs manual annotation. Our system, named Bayesian Hierarchical Disambiguator (BHD), uses the Internet, a ...

5 Special issue on Machine learning methods for text and images: Matching words and pictures

Kobus Barnard, Pinar Duygulu, David Forsyth, Nando de Freitas, David M. Blei, Michael I. Jordan

March 2003 **The Journal of Machine Learning Research**, Volume 3


Full text available:  [pdf\(789.04 KB\)](#) Additional Information: [full citation](#), [abstract](#), [citations](#), [index terms](#)

We present a new approach for modeling multi-modal data sets, focusing on the specific case of segmented images with associated text. Learning the joint distribution of image regions and words has many applications. We consider in detail predicting words associated with whole images (auto-annotation) and corresponding to particular image regions (region naming). Auto-annotation might help organize and access large collections of images. Region naming is a model of object recognition as a process ...

6 Text Extraction and Summarization: Combining multiple classifiers for text categorization

Khalid Al-Kofahi, Alex Tyrrell, Arun Vachher, Tim Travers, Peter Jackson

October 2001 **Proceedings of the tenth international conference on Information and knowledge management**

Full text available:  [pdf\(1.94 MB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#)

A major problem facing online information services is how to index and supplement large document collections with respect to a rich set of categories. We focus upon the routing of case law summaries to various secondary law volumes in which they should be cited. Given the large number (> 13,000) of closely related categories, this is a challenging task that is unlikely to succumb to a single algorithmic solution. Our fully implemented and recently deployed system shows that a superior classifica ...

Keywords: document classification, multi-classifier

7 Dialogue act modeling for automatic tagging and recognition of conversational speech

Andreas Stolcke, Noah Coccaro, Rebecca Bates, Paul Taylor, Carol Van Ess-Dykema, Klaus Ries, Elizabeth Shriberg, Daniel Jurafsky, Rachel Martin, Marie Meteer

September 2000 **Computational Linguistics**, Volume 26 Issue 3

Full text available:  [pdf\(2.53 MB\)](#)  Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#)
[Publisher Site](#)

We describe a statistical approach for modeling dialogue acts in conversational speech, i.e., speech-act-like units such as STATEMENT, QUESTION, BACKCHANNEL, AGREEMENT, DISAGREEMENT, and APOLOGY. Our model detects and predicts dialogue acts based on lexical, collocational, and prosodic cues, as well as on the discourse coherence of the

dialogue act sequence. The dialogue model is based on treating the discourse structure of a conversation as a hidden ...

8 Improving accuracy in word class tagging through the combination of machine learning systems

Hans van Halteren, Walter Daelemans, Jakub Zavrel
June 2001 **Computational Linguistics**, Volume 27 Issue 2

Full text available:  pdf(2.37 MB)  Additional Information: [full citation](#), [abstract](#), [references](#)
[Publisher Site](#)

We examine how differences in language models, learned by different data-driven systems performing the same NLP task, can be exploited to yield a higher accuracy than the best individual system. We do this by means of experiments involving the task of morphosyntactic word class tagging, on the basis of three different tagged corpora. Four well-known tagger generators (hidden Markov model, memory-based, transformation rules, and maximum entropy) are trained on the same corpus data. After comparison ...

9 Semantic annotation and integration: Towards the self-annotating web

Philipp Cimiano, Siegfried Handschuh, Steffen Staab
May 2004 **Proceedings of the 13th international conference on World Wide Web**

Full text available:  pdf(374.83 KB) Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

The success of the Semantic Web depends on the availability of ontologies as well as on the proliferation of web pages annotated with metadata conforming to these ontologies. Thus, a crucial question is where to acquire these metadata from. In this paper we propose PANKOW (Pattern-based Annotation through Knowledge on the Web), a method which employs an unsupervised, pattern-based approach to categorize instances with regard to an ontology. The approach is evaluated against the manual annotations ...

Keywords: information extraction, metadata, semantic annotation, semantic web

10 Hierarchical indexing and document matching in BoW

Maayan Geffet, Dror G. Feitelson
January 2001 **Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries**

Full text available:  pdf(434.24 KB) Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

BoW is an on-line bibliographical repository based on a hierarchical concept index to which entries are linked. Searching in the repository should therefore return matching topics from the hierarchy, rather than just a list of entries. Likewise, when new entries are inserted, a search for relevant topics to which they should be linked is required. We develop a vector-based algorithm that creates keyword vectors for the set of competing topics at each node in the hierarchy, and show how it works ...

11 Search improvement via automatic query reformulation


Susan Gauch, John B. Smith
July 1991 **ACM Transactions on Information Systems (TOIS)**, Volume 9 Issue 3

Full text available:  pdf(2.28 MB) Additional Information: [full citation](#), [references](#), [citations](#), [index terms](#), [review](#)

Keywords: Expert Systems, full-text information retrieval, online search assistance, query reformulation, textbases

12 Data mining: Mining the peanut gallery: opinion extraction and semantic classification of product reviews

Kushal Dave, Steve Lawrence, David M. Pennock
May 2003 **Proceedings of the twelfth international conference on World Wide Web**

Full text available:  [pdf\(327.95 KB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#)

The web contains a wealth of product reviews, but sifting through them is a daunting task. Ideally, an opinion mining tool would process a set of search results for a given item, generating a list of product attributes (quality, features, etc.) and aggregating opinions about each of them (poor, mixed, good). We begin by identifying the unique properties of this problem and develop a method for automatically distinguishing between positive and negative reviews. Our classifier draws on information ...

Keywords: document classification, opinion mining

13 [Integrating prosodic and lexical cues for automatic topic segmentation](#)

Gökhan TÜR, Andreas Stolcke, Dilek Hakkani-Tür, Elizabeth Shriberg

March 2001 **Computational Linguistics**, Volume 27 Issue 1

Full text available:  [pdf\(1.87 MB\)](#)  Additional Information: [full citation](#), [abstract](#), [references](#)
[Publisher Site](#)

We present a probabilistic model that uses both prosodic and lexical cues for the automatic segmentation of speech into topically coherent units. We propose two methods for combining lexical and prosodic information using hidden Markov models and decision trees. Lexical information is obtained from a speech recognizer, and prosodic features are extracted automatically from speech waveforms. We evaluate our approach on the Broadcast News corpus, using the DARPA-TDT evaluation metrics. Results show ...

14 [Technical best paper contest session: Multi-level annotation of natural scenes using dominant image components and semantic concepts](#)

Jianping Fan, Yuli Gao, Hangzai Luo

October 2004 **Proceedings of the 12th annual ACM international conference on Multimedia**

Full text available:  [pdf\(1.27 MB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Automatic image annotation is a promising solution to enable semantic image retrieval via keywords. In this paper, we propose a multi-level approach to annotate the semantics of **<i>natural scenes</i>** by using both the dominant image components (salient objects) and the relevant semantic concepts. To achieve automatic image annotation at the content level, we use salient objects as the dominant image components for image content representation and feature extraction. To ...

Keywords: adaptive EM algorithm, automatic image annotation, salient objects

15 [A model of multimedia information retrieval](#)

Carlo Meghini, Fabrizio Sebastiani, Umberto Straccia

September 2001 **Journal of the ACM (JACM)**, Volume 48 Issue 5

Full text available:  [pdf\(5.69 MB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#)

Research on multimedia information retrieval (MIR) has recently witnessed a booming interest. A prominent feature of this research trend is its simultaneous but independent materialization within several fields of computer science. The resulting richness of paradigms, methods and systems may, on the long run, result in a fragmentation of efforts and slow down progress. The primary goal of this study is to promote an integration of methods and techniques for MIR by contributing a conceptual model ...

Keywords: Description logics, fuzzy logics, multimedia information retrieval

16 Automatic generation of concise summaries of spoken dialogues in unrestricted domains

Klaus Zechner

September 2001

Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrievalFull text available:  [pdf\(209.67 KB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Automatic summarization of open domain spoken dialogues is a new research area. This paper introduces the task, the challenges involved, and presents an approach to obtain automatic extract summaries for multi-party dialogues of four different genres, without any restriction on domain. We address the following issues which are intrinsic to spoken dialogue summarization and typically can be ignored when summarizing written text such as newswire data: (i) detection and removal of speech disfl ...

17 Question-answering by predictive annotation

John Prager, Eric Brown, Anni Coden, Dragomir Radev

July 2000

Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrievalFull text available:  [pdf\(794.29 KB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#)

We present a new technique for question answering called Predictive Annotation. Predictive Annotation identifies potential answers to questions in text, annotates them accordingly and indexes them. This technique, along with a complementary analysis of questions, passage-level ranking and answer selection, produces a system effective at answering natural-language fact-seeking questions posed against large document collections. Experimental results show the effects of different parameter setti ...

18 RCV1: A New Benchmark Collection for Text Categorization Research

David D. Lewis, Yiming Yang, Tony G. Rose, Fan Li


August 2004 **The Journal of Machine Learning Research**, Volume 5Full text available:  [pdf\(628.29 KB\)](#) Additional Information: [full citation](#), [abstract](#), [citations](#), [index terms](#)

Reuters Corpus Volume I (RCV1) is an archive of over 800,000 manually categorized newswire stories recently made available by Reuters, Ltd. for research purposes. Use of this data for research on text categorization requires a detailed understanding of the real world constraints under which the data was produced. Drawing on interviews with Reuters personnel and access to Reuters documentation, we describe the coding policy and quality control procedures used in producing the RCV1 data, the inten ...

19 Term clustering of syntactic phrases

D. D. Lewis, W. B. Croft

December 1989

Proceedings of the 13th annual international ACM SIGIR conference on Research and development in information retrievalFull text available:  [pdf\(1.62 MB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#)

Term clustering and syntactic phrase formation are methods for transforming natural language text. Both have had only mixed success as strategies for improving the quality of text representations for document retrieval. Since the strengths of these methods are complementary, we have explored combining them to produce superior representations. In this paper we discuss our implementation of a syntactic phrase generator, as well as our preliminary experiments with producing phrase clusters. Th ...

20 Establishing the semantic web 1: SemTag and seeker: bootstrapping the semantic web via automated semantic annotation

Stephen Dill, Nadav Eiron, David Gibson, Daniel Gruhl, R. Guha, Anant Jhingran, Tapas Kanungo, Sridhar Rajagopalan, Andrew Tomkins, John A. Tomlin, Jason Y. Zien

May 2003 **Proceedings of the twelfth international conference on World Wide Web**Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#)

Full text available:  pdf(178.36 KB)

[terms](#)

This paper describes Seeker, a platform for large-scale text analytics, and SemTag, an application written on the platform to perform automated semantic tagging of large corpora. We apply SemTag to a collection of approximately 264 million web pages, and generate approximately 434 million automatically disambiguated semantic tags, published to the web as a label bureau providing metadata regarding the 434 million annotations. To our knowledge, this is the largest scale semantic tagging effort to ...

Keywords: automated semantic tagging, data mining, information retrieval, large text datasets, text analytics

Results 1 - 20 of 200

Result page: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [next](#)

The ACM Portal is published by the Association for Computing Machinery. Copyright © 2005 ACM, Inc.

[Terms of Usage](#) [Privacy Policy](#) [Code of Ethics](#) [Contact Us](#)

Useful downloads:  [Adobe Acrobat](#)  [QuickTime](#)  [Windows Media Player](#)  [Real Player](#)



US Patent & Trademark Office

[Subscribe \(Full Service\)](#) [Register \(Limited Service, Free\)](#) [Login](#)

Search: ☒ The ACM Digital Library ☐ The Guide

[classif*](#) and [automatic*](#) and [annotation*](#) and [concordanc*](#) and

SEARCH



[Feedback](#) [Report a problem](#) [Satisfaction survey](#)

Terms used

[classif](#) and [automatic](#) and [annotation](#) and [concordanc](#) and [normaliz](#) and [weight](#) and [word](#)

Found 20,783 of

150,138

Sort results
by

relevance



[Save results to a Binder](#)

Try an [Advanced Search](#)

Display
results

expanded form



[Search Tips](#)

Try this search in [The ACM Guide](#)

☐ Open results in a new
window

Results 21 - 40 of 200

Result page: [previous](#) [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [next](#)

Best 200 shown

Relevance scale ☐ ☐ ☐ ☐ ☐

21 [Computer Evaluation of Indexing and Text Processing](#)

G. Salton, M. E. Lesk

January 1968 **Journal of the ACM (JACM)**, Volume 15 Issue 1

Full text available: [pdf\(2.19 MB\)](#)

Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#)

Automatic indexing methods are evaluated and design criteria for modern information systems are derived.

22 [Modelling experience: Computational models for experiences in the arts, and multimedia](#)

Harini Sridharan, Hari Sundaram, Thanassis Rikakis

November 2003 **Proceedings of the 2003 ACM SIGMM workshop on Experiential telepresence**

Full text available: [pdf\(601.92 KB\)](#)

Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

In this paper, we develop formal computational models for three aspects of experiential systems for browsing media -- (a) context (b) interactivity through hyper-mediation and (c) context evolution using a memory model. Experiential systems deal with the problem of developing context adaptive mechanisms for knowledge acquisition and insight. Context is modeled as a union of graphs whose nodes represent concepts and where the edges represent the semantic relationships. The system context is the u ...

Keywords: art, context, experiential models, hyper-mediation, implications, semantic nets, wordNet

23 [Clustering: Probabilistic combination of text classifiers using reliability indicators: models and results](#)

Paul N. Bennett, Susan T. Dumais, Eric Horvitz

August 2002 **Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval**

Full text available: [pdf\(126.99 KB\)](#)

Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

The intuition that different text classifiers behave in qualitatively different ways has long motivated attempts to build a better metaclassifier via some combination of classifiers. We introduce a probabilistic method for combining classifiers that considers the context-sensitive reliabilities of contributing classifiers. The method harnesses *reliability indicators*---variables that provide a valuable signal about the performance of classifiers in different situations. We provide backgrou ...

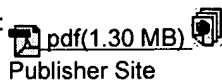
Keywords: classifier combination, metaclassifiers, reliability indicators, text classification

24 Automatic text categorization in terms of genre and author

Efstathios Stamatatos, George Kokkinakis, Nikos Fakotakis

December 2000 **Computational Linguistics**, Volume 26 Issue 4

Full text available:



Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#)

[Publisher Site](#)

The two main factors that characterize a text are its content and its style, and both can be used as a means of categorization. In this paper we present an approach to text categorization in terms of genre and author for Modern Greek. In contrast to previous stylometric approaches, we attempt to take full advantage of existing natural language processing (NLP) tools. To this end, we propose a set of style markers including analysis-level measures that represent the way in which the input text ha ...

25 Automatically extracting highlights for TV Baseball programs

Yong Rui, Anoop Gupta, Alex Acero

October 2000 **Proceedings of the eighth ACM international conference on Multimedia**

Full text available: [pdf\(1.08 MB\)](#)

Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#)

In today's fast-paced world, while the number of channels of television programming available is increasing rapidly, the time available to watch them remains the same or is decreasing. Users desire the capability to watch the programs time-shifted (on-demand) and/or to watch just the highlights to save time. In this paper we explore how to provide for the latter capability, that is the ability to extract highlights automatically, so that viewing time can be reduced.

We focus on the sp ...

Keywords: audio, baseball, highlights, summarization, television, video

26 Multimodal communication: Multimodal model integration for sentence unit detection

Mary P. Harper, Elizabeth Shriberg

October 2004 **Proceedings of the 6th international conference on Multimodal interfaces**

Full text available: [pdf\(469.02 KB\)](#)

Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

In this paper, we adopt a direct modeling approach to utilize conversational gesture cues in detecting sentence boundaries, called SUs, in video taped conversations. We treat the detection of SUs as a classification task such that for each inter-word boundary, the classifier decides whether there is an SU boundary or not. In addition to gesture cues, we also utilize prosody and lexical knowledge sources. In this first investigation, we find that gesture features complement the prosodic and le ...

Keywords: dialog, gesture, language models, multimodal fusion, prosody, sentence boundary detection

27 Text categorization: Text categorization by boosting automatically extracted concepts

Lijuan Cai, Thomas Hofmann

July 2003 **Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval**

Full text available: [pdf\(237.53 KB\)](#)

Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Term-based representations of documents have found wide-spread use in information retrieval. However, one of the main shortcomings of such methods is that they largely disregard lexical semantics and, as a consequence, are not sufficiently robust with respect to variations in word usage. In this paper we investigate the use of concept-based document

representations to supplement word- or phrase-based features. The utilized concepts are automatically extracted from documents via probabilistic latent ...

Keywords: boosting, classification, concept extraction, document categorization, lexical semantics, machine learning

28 Special issue on word sense disambiguation: Using corpus statistics and WordNet relations for sense identification

Claudia Leacock, George A. Miller, Martin Chodorow
March 1998 **Computational Linguistics**, Volume 24 Issue 1

Full text available:  [pdf\(1.35 MB\)](#)  Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#)
[Publisher Site](#)

Corpus-based approaches to word sense identification have flexibility and generality but suffer from a knowledge acquisition bottleneck. We show how knowledge-based techniques can be used to open the bottleneck by automatically locating training corpora. We describe a statistical classifier that combines topical context with local cues to identify a word sense. The classifier is used to disambiguate a noun, a verb, and an adjective. A knowledge base in the form of WordNet's lexical relations is ...

29 From reading to retrieval: freeform ink annotations as queries

Gene Golovchinsky, Morgan N. Price, Bill N. Schilit
August 1999 **Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval**

Full text available:  [pdf\(236.86 KB\)](#) Additional Information: [full citation](#), [references](#), [citations](#), [index terms](#)

Keywords: annotation-based queries, digital libraries, empirical evaluation, freeform digital ink, information appliances, information exploration, information retrieval, query expansion, query-mediated browsing, relevance feedback, user studies

30 Special issue on word sense disambiguation: Introduction to the special issue on word sense disambiguation: the state of the art

Nancy Ide, Jean Véronis
March 1998 **Computational Linguistics**, Volume 24 Issue 1

Full text available:  [pdf\(3.44 MB\)](#)  Additional Information: [full citation](#), [references](#), [citations](#)
[Publisher Site](#)

31 Finding factors: learning to classify case opinions under abstract fact categories

Stefanie Brünighaus, Kevin D. Ashley
June 1997 **Proceedings of the sixth international conference on Artificial intelligence and law**

Full text available:  [pdf\(1.27 MB\)](#) Additional Information: [full citation](#), [references](#), [citations](#), [index terms](#)

32 Local Feedback in Full-Text Retrieval Systems

R. Attar, A. S. Fraenkel
July 1977 **Journal of the ACM (JACM)**, Volume 24 Issue 3

Full text available:  [pdf\(1.36 MB\)](#) Additional Information: [full citation](#), [references](#), [citations](#), [index terms](#)

New Techniques for Disambiguation in Natural Language and Their Application to Biological Text

Filip Ginter, Jorma Boberg, Jouni Järvinen, Tapio Salakoski
August 2004 **The Journal of Machine Learning Research**, Volume 5

Full text available:  pdf(208.77 KB) Additional Information: [full citation](#), [abstract](#), [index terms](#)

We study the problems of disambiguation in natural language, focusing on the problem of gene vs. protein name disambiguation in biological text and also considering the problem of context-sensitive spelling error correction. We introduce a new family of classifiers based on ordering and weighting the feature vectors obtained from word counts and word co-occurrence in the text, and inspect several concrete classifiers from this family. We obtain the most accurate prediction when weighting by posi ...

34 Special issue on word sense disambiguation: Automatic word sense discrimination

Hinrich Schütze
March 1998 **Computational Linguistics**, Volume 24 Issue 1

Full text available:  pdf(1.97 MB)  Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#)
[Publisher Site](#)

This paper presents context-group discrimination, a disambiguation algorithm based on clustering. Senses are interpreted as groups (or clusters) of similar contexts of the ambiguous word. Words, contexts, and senses are represented in Word Space, a high-dimensional, real-valued space in which closeness corresponds to semantic similarity. Similarity in Word Space is based on second-order co-occurrence: two tokens (or contexts) of the ambiguous word are assigned to the same sense cluster if the wo ...

35 An automatic hierarchical image classification scheme

Jing Huang, S. Ravi Kumar, Ramin Zabih
September 1998 **Proceedings of the sixth ACM international conference on Multimedia**

Full text available:  pdf(855.42 KB) Additional Information: [full citation](#), [references](#), [citations](#), [index terms](#)

36 Information extraction: Unsupervised learning of soft patterns for generating definitions from online news

Hang Cui, Min-Yen Kan, Tat-Seng Chua
May 2004 **Proceedings of the 13th international conference on World Wide Web**


Full text available:  pdf(165.57 KB) Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Breaking news often contains timely definitions and descriptions of current terms, organizations and personalities. We utilize such web sources to construct definitions for such terms. Previous work has identified definitions using hand-crafted rules or supervised learning that constructs rigid, hard text patterns. In contrast, we demonstrate a new approach that uses flexible, soft matching patterns to characterize definition sentences. Our soft patterns are able to effectively accommodate the d ...

Keywords: definition generation, definitional question answering, pseudo-relevance feedback, soft patterns, unsupervised learning

37 Industry track papers: Learning domain-independent string transformation weights for high accuracy object identification

Sheila Tejada, Craig A. Knoblock, Steven Minton
July 2002 **Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining**

Full text available:  pdf(1.12 MB) Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#)

The task of object identification occurs when integrating information from multiple websites.

The same data objects can exist in inconsistent text formats across sites, making it difficult to identify matching objects using exact text match. Previous methods of object identification have required manual construction of domain-specific string transformations or manual setting of general transformation parameter weights for recognizing format inconsistencies. This manual process can be time consuming ...

38 Semantic web services: Meteor-s web service annotation framework

Abhijit A. Patil, Swapna A. Oundhakar, Amit P. Sheth, Kunal Verma

May 2004 **Proceedings of the 13th international conference on World Wide Web**

Full text available:  pdf(1.10 MB) Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

The World Wide Web is emerging not only as an infrastructure for data, but also for a broader variety of resources that are increasingly being made available as Web services. Relevant current standards like UDDI, WSDL, and SOAP are in their fledgling years and form the basis of making Web services a workable and broadly adopted technology. However, realizing the fuller scope of the promise of Web services and associated service oriented architecture will require further technological advances in ...

Keywords: ontology, semantic annotation of web services, semantic web services, web services discovery, wsd

39 Document analysis: Visual signature based identification of Low-resolution document images

Ardhendu Behera, Denis Lalanne, Rolf Ingold

October 2004 **Proceedings of the 2004 ACM symposium on Document engineering**

Full text available:  pdf(2.00 MB) Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

In this paper, we present (a) a method for identifying documents captured from low-resolution devices such as web-cams, digital cameras or mobile phones and (b) a technique for extracting their textual content without performing OCR. The first method associates a hierarchically structured visual signature to the low-resolution document image and further matches it with the visual signatures of the original high-resolution document images, stored in PDF form in a repository. The matching algorithm ...

Keywords: document visual signature, document-based meeting retrieval, documents' content extraction, low-resolution document image identification

40 Student best paper contest: Confidence-based dynamic ensemble for image annotation and semantics discovery

Beitao Li, Kingshy Goh

November 2003 **Proceedings of the eleventh ACM international conference on Multimedia**

Full text available:  pdf(275.37 KB) Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Providing accurate and scalable solutions to map low-level perceptual features to high-level semantics is critical for multimedia information organization and retrieval. In this paper, we propose a confidence-based dynamic ensemble (CDE) to overcome the shortcomings of the traditional *static* classifiers. In contrast to the traditional models, CDE can make dynamic adjustments to accommodate new semantics, to assist the discovery of useful low-level features, and to improve class-prediction ...

Results 21 - 40 of 200

Result page: [previous](#) [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [next](#)

The ACM Portal is published by the Association for Computing Machinery. Copyright © 2005 ACM, Inc.

[Terms of Usage](#) [Privacy Policy](#) [Code of Ethics](#) [Contact Us](#)

Useful downloads:  [Adobe Acrobat](#)  [QuickTime](#)  [Windows Media Player](#)  [Real Player](#)



US Patent & Trademark Office

[Subscribe \(Full Service\)](#) [Register \(Limited Service, Free\)](#) [Login](#)

 Search: ☒ The ACM Digital Library ☐ The Guide

classif* and automatic* and annotation* and concordanc* and

SEARCH


[Feedback](#) [Report a problem](#) [Satisfaction survey](#)

Terms used

classif and automatic and annotation and concordanc and normaliz and weight and word

Found 20,783 of

150,138

Sort results by

relevance

[Save results to a Binder](#)Try an [Advanced Search](#)

Display results

expanded form

[Search Tips](#)Try this search in [The ACM Guide](#)
☐ Open results in a new window
Results 41 - 60 of 200 Result page: [previous](#) [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [next](#)

Best 200 shown

Relevance scale ☐ ☐ ☐ ☐ ☐**41 Poster papers: Incremental context mining for adaptive document classification**

Rey-Long Liu, Yun-Ling Lu

 July 2002 **Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining**

 Full text available: [pdf\(641.82 KB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Automatic document classification (DC) is essential for the management of information and knowledge. This paper explores two practical issues in DC: (1) each document has its *context* of discussion, and (2) both the content and vocabulary of the document database is *intrinsically evolving*. The issues call for *adaptive document classification* (ADC) that adapts a DC system to the evolving contextual requirement of each document category, so that input documents may be classifie ...

Keywords: adaptive document classification, context text mining, incremental mining
42 Multidocument summarization: An added value to clustering in interactive retrieval

Manuel J. Maña-López, Manuel De Buenaga, José M. Gómez-Hidalgo

 April 2004 **ACM Transactions on Information Systems (TOIS)**, Volume 22 Issue 2

 Full text available: [pdf\(199.91 KB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

A more and more generalized problem in effective information access is the presence in the same corpus of multiple documents that contain similar information. Generally, users may be interested in locating, for a topic addressed by a group of similar documents, one or several particular aspects. This kind of task, called instance or aspectual retrieval, has been explored in several TREC Interactive Tracks. In this article, we propose in addition to the classification capacity of clustering techn ...

Keywords: Multidocument summarization, topic segmentation
43 The interaction of knowledge sources in word sense disambiguation

Mark Stevenson, Yorick Wilks

 September 2001 **Computational Linguistics**, Volume 27 Issue 3

 Full text available: [pdf\(2.16 MB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#)
[Publisher Site](#)

Word sense disambiguation (WSD) is a computational linguistics task likely to benefit from the tradition of combining different knowledge sources in artificial intelligence research. An important step in the exploration of this hypothesis is to determine which linguistic knowledge sources are most useful and whether their combination leads to improved

results. We present a sense tagger which uses several knowledge sources. Tested accuracy exceeds 94% on our evaluation corpus. Our system attempts ...

44 Biclustering Algorithms for Biological Data Analysis: A Survey

Sara C. Madeira, Arlindo L. Oliveira

January 2004 **IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)**, Volume 1 Issue 1

Full text available:  [pdf\(1.28 MB\)](#) Additional Information: [full citation](#)

Keywords: Biclustering, simultaneous clustering, coclustering, subspace clustering, bidimensional clustering, direct clustering, block clustering, two-way clustering, two-mode clustering, two-sided clustering, microarray data analysis, biological data analysis, gene expression data.

45 Towards the automatic identification of adjectival scales: clustering adjectives according to meaning

Vasileios Hatzivassiloglou, Kathleen R. McKeown

June 1993 **Proceedings of the 31st conference on Association for Computational Linguistics**

Full text available:  [pdf\(1.04 MB\)](#)  Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#)
[Publisher Site](#)

In this paper we present a method to group adjectives according to their meaning, as a first step towards the automatic identification of adjectival scales. We discuss the properties of adjectival scales and of groups of semantically related adjectives and how they imply sources of linguistic knowledge in text corpora. We describe how our system exploits this linguistic knowledge to compute a measure of similarity between two adjectives, using statistical techniques and without having access to ...

46 Artificial intelligence #1: SA_MetaMatch: relevant document discovery through document metadata and indexing

Hui S. Yau, J. Scott Hawker

April 2004 **Proceedings of the 42nd annual Southeast regional conference**

Full text available:  [pdf\(429.27 KB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

SA_MetaMatch, a component of the Standards Advisor (SA), is designed to find relevant documents through matching indices of metadata and document content. The elements in the metadata schema are mainly adopted from the Dublin Core (DC). The implementation of the XML metadata schema and coding follows the DC recommended guidelines. After metadata is generated manually for an unstructured document, or is extracted automatically from documents of well defined layout, they are stored in metadata file ...

Keywords: Dublin Core, document matching, index, metadata

47 Technical session 5: student best paper contest: LyricAlly: automatic synchronization of acoustic musical signals and textual lyrics

Ye Wang, Min-Yen Kan, Tin Lay Nwe, Arun Shenoy, Jun Yin

October 2004 **Proceedings of the 12th annual ACM international conference on Multimedia**

Full text available:  [pdf\(485.10 KB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

We present a prototype that automatically aligns acoustic musical signals with their corresponding textual lyrics, in a manner similar to manually-aligned karaoke. We tackle this problem using a multimodal approach, where the appropriate pairing of audio and text processing helps create a more accurate system. Our audio processing technique uses a combination of top-down and bottom-up approaches, combining the strength of low-level

audio features and high-level musical knowledge to determine ...

Keywords: audio/text synergy, karaoke, lyric alignment, music knowledge, vocal detection

48 Web mining and clustering: Stylistic and lexical co-training for web block classification

Chee How Lee, Min-Yen Kan, Sandra Lai

November 2004 **Proceedings of the 6th annual ACM international workshop on Web information and data management**

Full text available:  [pdf\(349.91 KB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Many applications which use web data extract information from a limited number of regions on a web page. As such, web page division into blocks and the subsequent block classification have become a preprocessing step. We introduce PARCELS, an open-source, co-trained approach that performs classification based on separate stylistic and lexical views of the web page. Unlike previous work, PARCELS performs classification on fine-grained blocks. In addition to table-based layout, the system handles ...

Keywords: PARCELS, co-training, lexical and stylistic learners, web page block classification, web page division

49 Image Categorization by Learning and Reasoning with Regions

Yixin Chen, James Z. Wang

August 2004 **The Journal of Machine Learning Research**, Volume 5

Full text available:  [pdf\(1.31 MB\)](#) Additional Information: [full citation](#), [abstract](#)

Designing computer programs to automatically categorize images using low-level features is a challenging research topic in computer vision. In this paper, we present a new learning technique, which extends Multiple-Instance Learning (MIL), and its application to the problem of region-based image categorization. Images are viewed as bags, each of which contains a number of instances corresponding to regions obtained from image segmentation. The standard MIL problem assumes that a bag is labeled p ...

50 Item-based top-N recommendation algorithms

Mukund Deshpande, George Karypis

January 2004 **ACM Transactions on Information Systems (TOIS)**, Volume 22 Issue 1

Full text available:  [pdf\(240.61 KB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

The explosive growth of the world-wide-web and the emergence of e-commerce has led to the development of *recommender systems*---a personalized information filtering technology used to identify a set of items that will be of interest to a certain user. User-based collaborative filtering is the most successful technology for building recommender systems to date and is extensively used in many commercial recommender systems. Unfortunately, the computational complexity of these methods grows l ...

Keywords: e-commerce, predicting user behavior, world wide web

51 Sources of Success for Boosted Wrapper Induction

David Kauchak, Joseph Smarr, Charles Elkan

August 2004 **The Journal of Machine Learning Research**, Volume 5

Full text available:  [pdf\(281.46 KB\)](#) Additional Information: [full citation](#), [abstract](#), [index terms](#)

In this paper, we examine an important recent rule-based information extraction (IE) technique named Boosted Wrapper Induction (BWI) by conducting experiments on a wider variety of tasks than previously studied, including tasks using several collections of natural text documents. We investigate systematically how each algorithmic component of BWI, in particular boosting, contributes to its success. We show that the benefit of boosting arises from the ability to reweight examples to learn specifi ...

52 Distributional clustering of English words

Fernando Pereira, Naftali Tishby, Lillian Lee

June 1993 **Proceedings of the 31st conference on Association for Computational Linguistics**

Full text available:  pdf(756.61 KB)

Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#)


 Publisher Site

We describe and evaluate experimentally a method for clustering words according to their distribution in particular syntactic contexts. Words are represented by the relative frequency distributions of contexts in which they appear, and relative entropy between those distributions is used as the similarity measure for clustering. Clusters are represented by average context distributions derived from the given words according to their probabilities of cluster membership. In many cases, the cluster ...

53 The automatic construction of large-scale corpora for summarization research

Daniel Marcu

August 1999 **Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval**

Full text available:  pdf(122.54 KB)

Additional Information: [full citation](#), [references](#), [citations](#), [index terms](#)

54 Semantic indexing for a complete subject discipline

Yi-Ming Chung, Qin He, Kevin Powell, Bruce Schatz

August 1999 **Proceedings of the fourth ACM conference on Digital libraries**

Full text available:  pdf(256.74 KB)

Additional Information: [full citation](#), [references](#), [index terms](#)

Keywords: MEDLINE, MEDSPACE, concept space, interspace, medical informatics, scalable semantics, semantic indexing, semantic retrieval

55 Topic detection and tracking in English and Chinese

Charles L. Wayne

November 2000 **Proceedings of the fifth international workshop on on Information retrieval with Asian languages**

Full text available:  pdf(688.09 KB)

Additional Information: [full citation](#), [abstract](#), [references](#)

Topic Detection and Tracking (TDT) refers to automatic techniques for discovering, threading, and retrieving topically related material in streams of data. Newswire and broadcast news are the canonical sources. In 1999, TDT research was extended from English to Chinese, and carefully annotated multilingual corpora were created. Researchers devised clever approaches to the cross-language challenge, and formal performance evaluations yielded very promising results. This paper outlines the 1999 ...

Keywords: detection, segmentation, speech, text, topic, tracking

56 Research track: SEWeP: using site semantics and a taxonomy to enhance the Web personalization process

M. Eirinaki, M. Vazirgiannis, I. Varlamis

August 2003 **Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining**

Full text available:  pdf(429.65 KB)

Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Web personalization is the process of customizing a Web site to the needs of each specific user or set of users, taking advantage of the knowledge acquired through the analysis of the user's navigational behavior. Integrating usage data with content, structure or user profile


data enhances the results of the personalization process. In this paper, we present SEWeP, a system that makes use of both the usage logs and the semantics of a Web site's content in order to personalize it. Web content is ...

Keywords: Web mining, Web personalization, concept hierarchies, semantic annotation of Web content

57 The Computer in the Humanities and Fine Arts

Sally Yeates Sedelow

June 1970 **ACM Computing Surveys (CSUR)**, Volume 2 Issue 2

Full text available:  pdf(2.01 MB)

Additional Information: [full citation](#), [references](#), [citations](#), [index terms](#)

58 Word association norms, mutual information, and lexicography

Kenneth Ward Church, Patrick Hanks

March 1990 **Computational Linguistics**, Volume 16 Issue 1

Full text available:  pdf(856.69 KB)

Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#)

 [Publisher Site](#)

The term *word association* is used in a very particular sense in the psycholinguistic literature. (Generally speaking, subjects respond quicker than normal to the word *nurse* if it follows a highly associated word such as *doctor*.) We will extend the term to provide the basis for a statistical description of a variety of interesting linguistic phenomena, ranging from semantic relations of the doctor/nurse type (content word/content word) to lexico-syntactic co-occurrence const ...

59 The SMART lab report

Mike Lesk, Donna Harman, Edward A. Fox, Harry Wu, Chris Buckley

April 1997 **ACM SIGIR Forum**, Volume 31 Issue 1


Full text available:  pdf(1.65 MB)

Additional Information: [full citation](#), [index terms](#)

60 Industry track papers: On the potential of domain literature for clustering and Bayesian network learning

Peter Antal, Patrick Glenisson, Geert Fannes

July 2002 **Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining**

Full text available:  pdf(1.10 MB)

Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Thanks to its increasing availability, electronic literature can now be a major source of information when developing complex statistical models where data is scarce or contains much noise. This raises the question of how to integrate information from domain literature with statistical data. Because quantifying similarities or dependencies between variables is a basic building block in knowledge discovery, we consider here the following question. Which vector representations of text and which st ...

Keywords: Bayesian networks, clustering, data mining, text mining

Results 41 - 60 of 200

Result page: [previous](#) [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [next](#)

The ACM Portal is published by the Association for Computing Machinery. Copyright © 2005 ACM, Inc.

[Terms of Usage](#) [Privacy Policy](#) [Code of Ethics](#) [Contact Us](#)

Useful downloads:  [Adobe Acrobat](#)  [QuickTime](#)  [Windows Media Player](#)  [Real Player](#)

Patent Assignment Abstract of Title

Total Assignments: 1**Application #:** 09563900**Filing Dt:** 05/03/2000**Patent #:** NONE**Issue Dt:****PCT #:** NONE**Publication #:** NONE**Pub Dt:****Inventor:** Christopher Clayton McConnell**Title:** METHODS, APPARATUS, AND DATA STRUCTURES FOR ANNOTATING A DATABASE DESIGN SCHEMA
AND/OR INDEXING ANNOTATIONS**Assignment: 1****Reel/Frame:** 011040/0828 **Received:** 09/07/2000 **Recorded:** 08/28/2000 **Mailed:** 10/20/2000 **Pages:****Conveyance:** ASSIGNMENT OF ASSIGNORS INTEREST (SEE DOCUMENT FOR DETAILS).**Assignor:** MCCONNELL, CHRISTOPHER CLAYTON**Exec Dt:** 08/23/2000**Assignee:** MICROSOFT CORPORATION

ONE MICROSOFT WAY

REDMOND, WASHINGTON 98052

Correspondent: STRAUB & POKOTYLO

JOHN C. POKOTYLO, ESQ.

1 BETHANY ROAD

SUITE 83, BUILDING 6

HAZLET, NJ 07730

Search Results as of: 2/9/2005 4:24:59

If you have any comments or questions concerning the data displayed, contact OPR / Assignments at 703-308-9723
Web interface last modified: Oct. 5, 2002